BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: G03G	A2	(11) International Publication Number:	WO 00/04422
		(43) International Publication Date:	27 January 2000 (27.01.2000)

(21) International Application Number: PCT/US99/16515

(22) International Filing Date: 19 July 1999 (19.07.1999)

(30) Priority Data:
09/120,023 17 July 1998 (17.07.1998) US

(60) Parent Application or Grant
LEIGHTON, F., Thomson [/]; (). LEIGHTON, F., Thomson [/]; (). JUDSON, David, H. ; ().

Published

(54) Title: METHOD FOR IMAGE PROCESSING TO FACILITATE COPY PROTECTION

(54) Titre: PROCEDE DE TRAITEMENT D'IMAGE FACILITANT LA PROTECTION CONTRE LA COPIE

(57) Abstract

A method for mapping a work (e.g., a picture, a document or the like) comprising a large number of bytes into a relatively small bit string to facilitate copy protection. The technique creates first string from the work with the property that if one modifies the work by some given degree (e.g., through intentional variation or perhaps merely through aging), a second string (generated by the inventive technique) and associated with the modified work is not meaningfully distinct from the first string. When applied through the algorithm, works that are similar result in the same or similar strings. As a result, an owner of the content sought to be protected may apply the algorithm to copies of the work for copy protection purposes.

(57) Abrégé

Procédé permettant d'effectuer le mappage d'une oeuvre (graphique, document, etc.) constituée d'un grand nombre d'octets de façon à obtenir une chaîne binaire relativement petite afin de faciliter la protection contre la copie. La technique consiste à créer, à partir de l'oeuvre, une première chaîne caractérisée par le fait que si quelqu'un modifie ladite oeuvre à un certain degré (modification intentionnelle ou simple vieillissement), une seconde chaîne (générée par la technique de l'invention) associée à l'oeuvre modifiée n'est pas sensiblement distincte de la première chaîne. Une fois que l'algorithme leur a été appliqué, des oeuvres similaires donnent des chaînes identiques ou similaires. Le propriétaire du contenu à protéger peut donc appliquer cet algorithme aux copies de l'oeuvre à des fins de protection contre la copie.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : G03G — A2

(11) International Publication Number: WO 00/04422

(43) International Publication Date: 27 January 2000 (27.01.00)

(21) International Application Number: PCT/US99/16515

(22) International Filing Date: 19 July 1999 (19.07.99)

(30) Priority Data:
09/120,023 17 July 1998 (17.07.98) US

(71)(72) Applicant and Inventor: LEIGHTON, F., Thomson [US/US]; 15 Charlesden Park, Newtonville, MA 02160 (US).

(74) Agent: JUDSON, David, H.; Hughes & Luce, L.L.P., Suite 2800, 1717 Main Street, Dallas, TX 75210 (US).

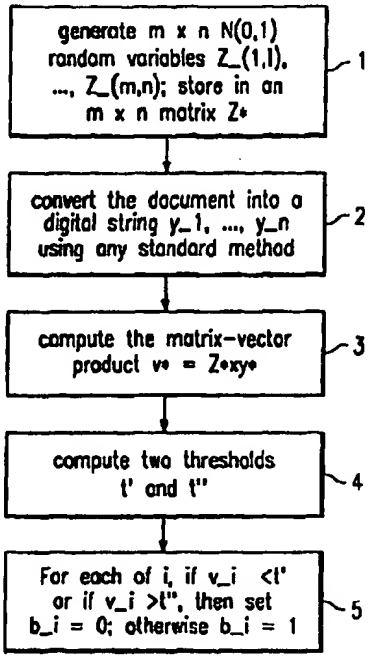
(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: METHOD FOR IMAGE PROCESSING TO FACILITATE COPY PROTECTION

(57) Abstract

A method for mapping a work (e.g., a picture, a document or the like) comprising a large number of bytes into a relatively small bit string to facilitate copy protection. The technique creates first string from the work with the property that if one modifies the work by some given degree (e.g., through intentional variation or perhaps merely through aging), a second string (generated by the inventive technique) and associated with the modified work is not meaningfully distinct from the first string. When applied through the algorithm, works that are similar result in the same or similar strings. As a result, an owner of the content sought to be protected may apply the algorithm to copies of the work for copy protection purposes.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Description

METHOD FOR IMAGE PROCESSING TO FACILITATE COPY PROTECTION

BACKGROUND OF THE INVENTION

Technical Field

The present invention relates generally to securing content against wrongful duplication and, more particularly, techniques for processing an image into a digital string so that images with similar features (e.g., illicit or cropped copies) get mapped into a digital string that is close to the original string.

Description of the Related Art

The proliferation of digitized media and the ease with which digital files can be copied especially over public computer networks (e.g., the Internet) has created a need for copy and copyright enforcement schemes. Conventional cryptographic systems permit only valid keyholders access to encrypted data, but once such data is decrypted there is no way to track its reproduction or retransmission. Such schemes thus provide insufficient protection against unauthorized reproduction of information. It is known in the prior art to provide a so-called digital "watermark" on a document to address this problem. A "watermark" is a visible or preferably invisible identification code that is permanently embedded in the data and thus remains present within the data after any decryption process. One example of a digital watermark would be a visible "seal" placed over an image to identify the copyright owner. However, the watermark might also contain additional information, including the identity of the purchaser of a particular copy of the material.

Many schemes have been proposed for watermarking digital data. In a known watermarking procedure, each copy of a document D is varied slightly so as to look the same to the user but also so as to include the identity of the purchaser. The watermark consists of the variations that are unique to each copy. The idea behind such schemes is that the watermark should be hard to remove without destroying the document. Thus, a copy of a watermarked document should be traceable back to the specific version of the original from which it was created. Although certain watermarking techniques provide advantages, there remains a need for a more general approach to the problems associated with content protection.

SUMMARY OF THE INVENTION

It is a primary object of the present invention to provide an improved method for securing content against wrongful duplication.

It is still another object of this invention to provide improved methods for processing an image (or other work) into a digital string so that images with similar features (e.g., illicit or cropped copies) get mapped into a digital string that is close to the original digital string.

It is another more general object of this invention to protect against the wrongful proliferation of digitized media over public computer networks (e.g., the Internet).

A still further object of this invention is to provide for more efficient and reliable copy and copyright enforcement schemes.

The above as well as additional objects, features, and advantages of the present invention will become apparent in the following detailed written description.

BRIEF DESCRIPTION OF THE DRAWINGS

The novel features believed characteristic of the invention are set forth in the appended claims The invention itself however, as well as a preferred mode of use, further objects and advantages thereof, are best understood by reference to the following Detailed Description of an illustrative embodiment when read in conjunction with the accompanying Drawings, wherein:

Figure 1 is a flowchart of the inventive method.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention is preferably implemented in software (e.g., a series of program instructions) executed in a computer. It is assumed that the content to be protected in a picture or other image. The invention may be applied to any content (including, without limitation, video, graphics, sound, text, etc.) irrespective of its form, however. For purposes of illustration, the invention will be described in the context of a picture. The picture is digitized in any convenient manner.

According to the invention, once digitized, the picture is run through a hashing algorithm and reduced to approximately thirty (30) bits so that most any other picture that was a small modification of the original gets mapped to the same string (with maybe a couple of bits different). This is useful in watermarking since this string, then, could be used as part of x_1, ..., x_n when one uses H(x_1 ... x_n|Do not copy) as an offset watermark. When this approach is followed, the watermark used for different pictures is different (which makes it harder for the adversary to remove), but it is still possible for someone who knows H to regenerate the watermark. In particular, the watermark used for different pictures is different (which makes it harder for the adversary to remove) but it is still possible for someone who knows H to regenerate the watermark for a correlation test (without access to a database), since even if the

adversary changes the picture a little, he cannot change the string x_1, ..., x_n very much; so, it is possible to later regenerate the original x_1, ..., x_n by computing the derived values from the picture and trying all possibilities that are within a few bits.

Such a mapping function is also useful in Web search engines. For example, assume one wants to search for copies of a document/picture on the Web even if such copies are changed a little bit. One could use the present invention and compute the string for every candidate picture on the Web and see which ones were close to the string for the original. From this information, one could check the candidates that match by eye, thereby cutting the search space by a huge factor.

There may also be applications in the domain of organizing pictorial data into similarity classes, which could be done according to the digital string produced by the following algorithm.

Implementation

With reference now to the flowchart of Figure 1, the inventive mapping scheme has a number of steps.

Step 1: generate m x n N(0,1) random variables Z_(1,1), ..., Z_(m,n). Store these values in an m x n matrix Z*. These values will be used for all documents and will be kept secret. This may be done in an off-line process.

Step 2: convert the document into a digital string y_1, ..., y_n using any standard method. If possible, normalize the y_i values so that over the universe of all documents, y_i can be thought of as a N(0,1) random variable. (The latter step is not required, but it makes the process work better.) Treat the n values as a vector y*.

Step 3: compute the matrix-vector product v* = Z*xy*. Let v_1, ..., v_m denote the values of v*. Note that each v_i can be thought of as a normal random variable, no matter what the y values are. For example, the mean of each v_i is M = (y_1 + ... + y_m)/m.

Step 4: compute two thresholds t' and t" as follows. Select t' so that the probability that y_1 x u_1 + ... + y_m x u_m <t' is close to about 1/4 when each u_i is an independent N(0,1) random variable. Select t" so that the probability that y_1 x u_1 + ... + y_m x u_m > t" is close to about 1/4. There are many possible variations of this step, e.g., one with just one threshold at the value t =M, or by adjusting the two threshold values.

Step 5: For each of i, if v_i <t' or if v_i >t", then set b_i = 0. Otherwise, set b_i = 1. (Note that each b_i will be equally likely to be 0 or 1.) The values b_1, ..., b_m form the desired bit string for the document.

The mapping function reduces the content (to be protected) from a large number of bytes to a string having a small number of bits. Changing just a few values of y* (or all of the values of y* by a little bit), however, does not have a meaningful effect on the values of b* that are computed. For example, in order to change the value of b_j, the adversary must change enough of the y_i's by enough so that the corresponding v_j is pushed across a threshold. Without knowing the Z* values, this is hard for the adversary to do if he is restricted to making small changes in the y_i's (since most of the v_j's are not likely to be very near a threshold in the first place). In fact, there is a formal mathematical proof that says that the amount by which the adversary is required to change the y_i's to effect a significant change in the b_j's is the maximum possible for any scheme.

Thus, the mapping function (when applied to copies of the document) is quite useful for copy protection purposes. Generalizing, the technique described above creates a first string from a picture (or other work) with the property that if one modifies the picture by some given degree (e.g., through intentional variation or perhaps merely through aging), a second string (generated by the inventive technique) and associated with the modified picture is not meaningfully distinct from the first string. The invention can be used to reduce a picture (or other work) comprising a large number of bytes into a string having a small number of bits, and thus pictures that are similar result in the same or similar strings. As a result, an owner of the content sought to be protected may apply the algorithm to copies of the work for copy protection purposes.

As described above, aspects of the present invention pertain to specific "method steps" implementable on one or more computers. In an alternate embodiment, the invention may be implemented as a computer program product for use with a computer system. Those skilled in the art should readily appreciate that programs defining the functions of the present invention can be delivered to a computer in many forms, which include, but are not limited to:
(a) information permanently stored on non-writable storage media (e.g. read only memory devices within a computer such as ROM or optical disks readable by CD-ROM drive);
(b) information alterably stored on writable storage media (e.g., floppy disks within a diskette drive or a hard disk drive); or (c) information conveyed to a computer through communication media, such as through a computer or telephone or other network. It should be understood, therefore, that such media, when carrying computer readable instructions that direct the method functions of the present invention, represent alternate embodiments of the present invention.

In addition, although the various methods described are conveniently implemented in a general purpose computer selectively activated or reconfigured by software, one of ordinary skill in the art would also recognize that such methods may be carried out in hardware, in firmware, or in more specialized apparatus (e.g., a secure chip) constructed to perform the required method steps.

While the invention has been particularly shown and described with reference to a preferred embodiment, it will be understood by those skilled in the art that various changes in form and detail may be made therein without departing from the spirit and scope of the invention.

Claims

CLAIMS

1. A method for mapping a document comprising a large number of bytes into a bit string to facilitate copy protection of the document, comprising the steps of:

generating m x n N(0,1) random variables Z_(1,1), ..., Z_(m,n) and storing the random variables in an m x n matrix Z*;

processing the document into a digital string y_1, ..., y_n;

normalizing the y_i values to generate a vector y*;

computing a matrix-vector product v* = Z*xy*, where v_1, ..., v_m denote the values of v*;

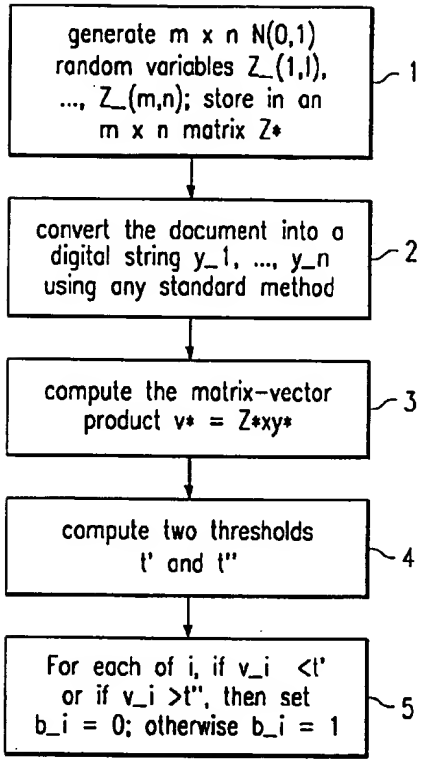
selecting first and second thresholds t' and t" where t' is selected so that the probability that y_1 x u_1 + ... + y_m x u_m <t' is close to about 1/4 when each u_i is an independent N(0,1) random variable, and where t" is selected so that the probability that y_1 x u_1 + ... + y_m x u_m > t" is close to about 1/4;

each i, determining if v_i <t' or if v_i >t"; and

if v_i is less than t' or if v_i is greater than t", then set b_i = 0;

if v_i is not less than t' or if v_i is not greater than t", set b_i = 1; and

concatenating the values b_1, ..., b_m into the bit string.


generate m x n N(0,1) random variables Z_(1,1), ..., Z_(m,n); store in an m x n matrix Z*
1
convert the document into a digital string y_1, ..., y_n using any standard method
2
compute the matrix-vector product v* = Z*xy*
3
compute two thresholds t' and t"
4
For each of i, if v_i <t' or if v_i >t", then set b_i = 0; otherwise b_i = 1
5